A Novel Auxiliary Agarolytic Pathway Expands Metabolic Versatility in the Agar-Degrading Marine Bacterium *Colwellia echini* A3^T

Duleepa Pathiraja,[a] Line Christiansen,[b*] Byeonghyeok Park,[a] Mikkel Schultz-Johansen,[b*] Geul Bang,[c,d] ⓘPeter Stougaard,[b*] ⓘIn-Geol Choi[a]

[a]Department of Biotechnology, College of Life Sciences and Biotechnology, Korea University, Seoul, South Korea
[b]Department of Plant and Environmental Sciences, University of Copenhagen, Copenhagen, Denmark
[c]Biomedical Omics Group, Korea Basic Science Institute, Chungbuk, South Korea
[d]College of Pharmacy, Korea University, Sejong, South Korea

Duleepa Pathiraja and Line Christiansen contributed equally to this work. Author order was determined by the contributions to analyzing data and writing the original manuscript.

ABSTRACT Marine microorganisms encode a complex repertoire of carbohydrate-active enzymes (CAZymes) for the catabolism of algal cell wall polysaccharides. While the core enzyme cascade for degrading agar is conserved across agarolytic marine bacteria, gain of novel metabolic functions can lead to the evolutionary expansion of the gene repertoire. Here, we describe how two less-abundant GH96 α-agarases harbored in the agar-specific polysaccharide utilization locus (PUL) of *Colwellia echini* strain A3^T facilitate the versatility of the agarolytic pathway. The cellular and molecular functions of the α-agarases examined by genomic, transcriptomic, and biochemical analyses revealed that α-agarases of *C. echini* A3^T create a novel auxiliary pathway. α-Agarases convert even-numbered neoagarooligosaccharides to odd-numbered agaro- and neoagarooligosaccharides, providing an alternative route for the depolymerization process in the agarolytic pathway. Comparative genomic analysis of agarolytic bacteria implied that the agarolytic gene repertoire in marine bacteria has been diversified during evolution, while the essential core agarolytic gene set has been conserved. The expansion of the agarolytic gene repertoire and novel hydrolytic functions, including the elucidated molecular functionality of α-agarase, promote metabolic versatility by channeling agar metabolism through different routes.

IMPORTANCE *Colwellia echini* A3^T is an example of how the gain of gene(s) can lead to the evolutionary expansion of agar-specific polysaccharide utilization loci (PUL). *C. echini* A3^T encodes two α-agarases in addition to the core β-agarolytic enzymes in its agarolytic PUL. Among the agar-degrading CAZymes identified so far, only a few α-agarases have been biochemically characterized. The molecular and biological functions of two α-agarases revealed that their unique hydrolytic pattern leads to the emergence of auxiliary agarolytic pathways. Through the combination of transcriptomic, genomic, and biochemical evidence, we elucidate the complete α-agarolytic pathway in *C. echini* A3^T. The addition of α-agarases to the agarolytic enzyme repertoire might allow marine agarolytic bacteria to increase competitive abilities through metabolic versatility.

KEYWORDS agarase, α-agarase, GH96, agar metabolism, polysaccharide utilization loci, gene gain, novel auxiliary pathway, metabolic versatility

Macroalgae contribute to the primary production in coastal and pelagic marine ecosystems and provide a rich nutrient source for marine microorganisms (1, 2). A significant part of the marine algal biomass comprises cell wall polysaccharides such as agar, porphyran, and carrageenan in red macroalgae, fucan in brown algae, and ulvan in green algae (3–5). Bioconversion of algal polysaccharides in marine ecosystems is carried out

by microorganisms that encode a repertoire of carbohydrate-active enzymes (CAZymes) for the efficient hydrolysis of various algal cell wall polysaccharides. This CAZyme repertoire is diverse and varies among marine polysaccharide-degrading microorganisms (6).

CAZyme genes are often clustered in genomic regions, known as polysaccharide utilization loci (PULs) (7–10). PULs were initially identified in *Bacteroidetes*, and similar genomic loci were identified later in other phyla (11, 12). Genes in the PULs encode canonical SusC/SusD proteins for binding and transportation of oligosaccharides (11, 12), CAZymes (including glycoside hydrolases), transporters, transcriptional regulators, and auxiliary enzymes (e.g., sulfatases and phosphatases) (10). The genes in PULs are often regulated as a unit (regulon) (7–9). The SusC/SusD pair is exclusively found within the *Bacteroidetes*, whereas the non-*Bacteroidetes* have TonB-dependent transporters instead of SusC/SusD (11, 12). While certain marine microorganisms share homologous glycoside hydrolases within the PULs, less conserved and functionally novel enzymes may also be found. The identification of coregulated genes within a PUL often leads to the characterization of novel or rare CAZymes and auxiliary hydrolytic pathways (13, 14).

Among different macroalgal cell wall polysaccharides, agar is a hetero-polymer composed of two distinct α-1,3 and β-1,4 glycosidic linkage groups between 3,6-anhydro-L-galactose (L-AHG) and D-galactose (15). α-Agarases and β-agarases catalyze the hydrolysis of two different glycosidic linkages. Recent studies suggest that agarolytic pathways in marine microorganisms are biased toward cleavage of the β-1,4 glycosidic linkage (16–19). Model agarolytic bacteria such as *Saccharophagus degradans* strain 2-40, *Pseudoalteromonas atlantica* strain T6c, *Vibrio* sp. EJY3, and *Zobellia galactanivorans* Dsij[T] only use β-agarases to hydrolyze agar polymers by β-1,4 glycosidic link cleaving (16, 20–22). β-Agarases have been well characterized and are classified into the glycoside hydrolase (GH) families GH16, GH50, GH86, and GH118 based on their sequence similarity (23). Unlike these well-characterized and ubiquitous β-agarases, α-agarases are rarely found in nature and cleave the α-1,3 glycosidic linkage in agar. Only a few α-agarases, such as those from *Alteromonas agarilyticus* strain GJ1B, *Thalassotalea agarivorans* strain JAMB-A33[T], and *Thalassomonas* sp. strain LD5 have previously been reported (24–26). All published α-agarases belong to a single CAZy family, GH96 (27). Another enzyme, GH117 α-neoagarobiose hydrolase (α-NABH), hydrolyzes α-1,3 glycosidic bonds but, in contrast to α-agarases, α-NABH exclusively releases L-AHG from the nonreducing end of neoagarooligosaccharides (NAOS) (18). The limited genetic diversity and the rare natural occurrence of α-agarases, compared to β-agarases, raise a fundamental question about the metabolic benefit and evolutionary niche created by α-agarases.

The agarolytic bacterium *Colwellia echini* A3[T], previously isolated from the microbiome of sea urchin (28), encodes both β-agarases and α-agarases organized in a PUL-like genomic locus (29). Here, we describe in detail the β- and α-agarolytic pathways in *C. echini* A3[T] for the hydrolysis of agar to monosaccharides, using genomics, transcriptomics, and biochemical characterizations of β- and α-agarases. We used the agarolytic pathway of *C. echini* A3[T] as an experimental model system for elucidating how gene gain contributes to the emergence of novel auxiliary metabolic pathways and diversification of the core β-agarolytic pathway. Comparative genomic analysis of 30 other marine bacteria that contain agarolytic enzymes supports that the agarolytic enzyme repertoire of marine bacteria has diversified during evolution while conserving essential core genes. This study provides new insights into the role of α-agarases in the enzymatic degradation of agar and maps the environmental distribution and conservation of α-agarolytic pathways in marine bacteria.

RESULTS

Genomic and transcriptomic analyses show that agar hydrolysis is encoded by a complex gene cluster in *Colwellia echini* A3[T]. *C. echini* A3[T] possesses a repertoire of enzymes capable of degrading complex marine polysaccharides (28). In this study, we focused on the repertoire of agarolytic enzymes produced by *C. echini* A3[T] and extensively studied their agar degradation potential. Whole-genome sequencing and annotation of *C. echini* A3[T] revealed a cluster of genes responsible for agar hydrolysis and
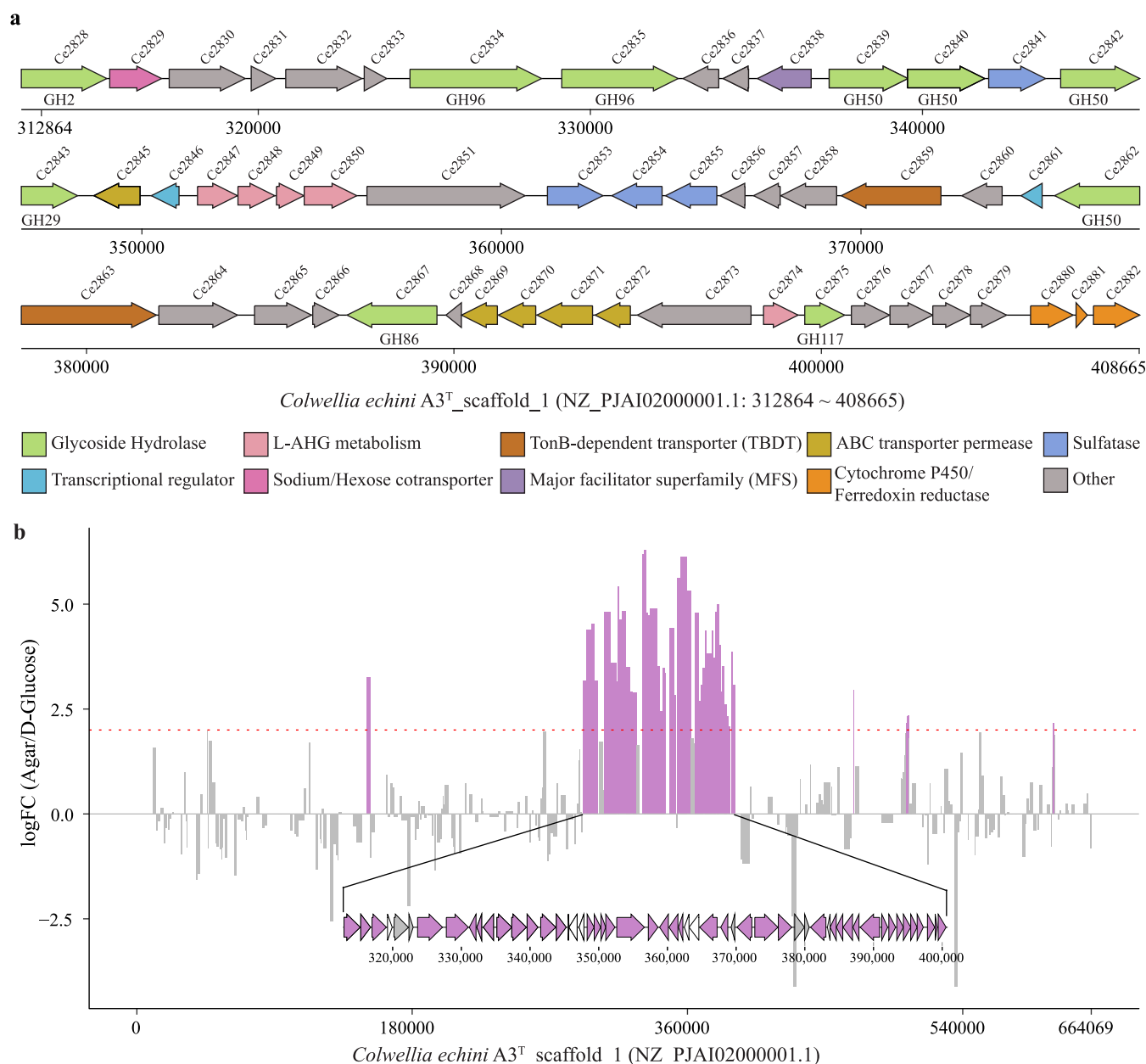
FIG 1 Genomic and transcriptomic analysis of the PUL involved in agar metabolism (AGA PUL) in *C. echini* A3[T]. (a) Genes in the AGA PUL were annotated according to their CAZy family or enzymatic activity. Nine main categories of genes (glycoside hydrolase, L-AHG metabolism, TonB-dependent transporters, ABC transporter permease, sulfatase, transcriptional regulators, sodium/hexose cotransporters, major facilitator superfamily, and cytochrome P450/ferredoxin reductase) were identified based on their function in agar metabolism. The names of the genes and their functions are listed in Table S1 in the supplemental material. (b) Log fold change (logFC) of the gene expression levels of the genes within scaffold 1 of the *C. echini* A3[T] genome (NCBI Reference Sequence accession NZ_PJAI02000001.1). Genes with logFC > 2 and $P < 0.05$ were considered differentially expressed. The gene expression profile of *C. echini* A3[T] was obtained from the RNA-seq analysis of cells grown in marine minimal medium containing either agar or D-glucose as the carbon source. Log fold changes (logFC) of the gene expression levels in agar compared to D-glucose are shown.

metabolism. The gene cluster (98.7 kb) contains various agar-specific CAZymes, TonB-dependent transporters (TBDT), transcriptional regulators, and auxiliary enzymes (sulfatases and phosphatases). The organization of this region has similarities to the PULs described in *Bacteroidetes* (10). Therefore, this gene cluster was assumed to be a PUL specific for agar (designated "AGA PUL") in *C. echini* A3[T] (29) (Fig. 1a) (Table S1 and Data S1 in the supplemental material).

The AGA PUL of *C. echini* A3[T] contains all the genes required for agar hydrolysis, uptake, metabolism, and transcriptional control. This comprises nine GH enzymes,

including two GH96 α-agarases (Ce2834 and Ce2835), one GH86 β-agarase (Ce2867), four GH50 β-agarases (Ce2839, Ce2840, Ce2842, and Ce2862), one GH2 agarolytic β-galactosidase (ABG) (Ce2828), and one GH117 α-neoagarobiose hydrolase (Ce2875). Four putative L-AHG metabolic enzymes are located next to each other in the AGA PUL; 3,6-anhydro-L-galactose dehydrogenase (AHG dehydrogenase) (Ce2850), 3,6-anhydro-L-galactonate cycloisomerase (AHGA cycloisomerase) (Ce2847), 2-keto-3-deoxy-L-galactonate 5-dehydrogenase (KDG dehydrogenase) (Ce2848), and 2,5-diketo-3-deoxy-L-galactonate 5-reductase (DDG reductase) (Ce2849). Another putative L-AHG metabolic enzyme, 2-keto-3-deoxy-D-gluconate kinase (KDG kinase) (Ce2874), is located within the PUL but away from the other four. The 2-dehydro-3-deoxy-6-phosphogalactonate aldolase (KDPG aldolase) gene is not found within the AGA PUL, and two putative aldolase genes are found elsewhere in the genome (Ce755 and Ce359). The AGA PUL also contains proteins that are expected to be involved in oligo- and monosaccharide transport, including TBDT (Ce2859 and Ce2863), ABC transporters (Ce2845, Ce2869, Ce2870, Ce2871, and Ce2872), a major facilitator superfamily (MFS) transporter (Ce2838), and a sodium/hexose cotransporter (Ce2829). Two transcriptional regulators (Ce2846 and Ce2861) are also located within the AGA PUL.

Transcriptome sequencing (RNA-seq) gene expression profiling was performed on *C. echini* A3^T grown in marine minimal medium containing agar as the carbon source and in D-glucose. The expression profile supported that genes in the coinciding AGA PUL region were strongly upregulated (log$_2$ fold change > 2) compared to other genomic regions (Fig. 1b). All the AGA PUL annotated genes were upregulated except for the transcriptional regulators (Table S2 and Data S2). Two L-AHG metabolic genes (Ce2848 and Ce2847) were the most upregulated in the PUL (log fold change [FC] > 6), followed by a TBDT (Ce2863), β-agarases (Ce2862, Ce2839, and Ce2867), and an α-agarase (Ce2834).

Phylogenetic analysis and molecular characterization of GH96 α-agarases. *C. echini* A3^T has two GH96 α-agarases (Ce2834 and Ce2835), each with $>70\%$ amino acid sequence identity to α-agarases from *Algibacillus agarilyticus* strain RQJ05^T (Table S3). In addition, Ce2834 has 74% amino acid identity to a previously characterized α-agarase from *Alteromonas agarilyticus* GJ1B (24) and Ce2835 has 71% amino acid identity to previously characterized α-agarase from *Thalassotalea agarivorans* JAMB-A33^T. We retrieved amino acid sequences of α-agarase homologs containing GH96 catalytic modules from the NCBI GenBank database (as of December 2020). The α-agarases were clustered into two distinct groups (group I and II) by sequence similarity (Fig. 2a, Fig. S1). The GH96 members of groups I and II taxonomically belong to the phyla *Proteobacteria* and *Bacteroidetes*, respectively. Sequence analysis revealed that, apart from a common GH96 catalytic domain, the modular architecture differed between proteins in group I and II (Fig. 2b). The α-agarases from group I contained conserved features as follows: (i) noncatalytic carbohydrate binding module family 6 (CBM6), also found in β-agarases (30); (ii) thrombospondin type 3 (TSP3) repeats, which are short aspartate-rich repeats proposed to serve as a continuous series of calcium binding sites (31); and (iii) PA14 modules upstream of the GH96 catalytic module. None of these above-mentioned modules were found in group II. The group II members contained a distinct T9SS sorting module downstream of the GH96 catalytic module (32) that was not found in group I. The biochemical activities of four α-agarases from group I were experimentally verified (24–26), but none of the group II α-agarases has been examined yet. We constructed a gene tree using the amino acid sequences of α-agarases from *Gammaproteobacteria* in group I (Fig. 2c, Fig. S2). All α-agarases formed a single clade, suggesting they are homologs found in closely related *Gammaproteobacteria*. The GH96 homologs of closely related *Gammaproteobacteria* show $>60\%$ amino acid identity with the α-agarases from *C. echini* A3^T. Few ortholog/paralog relationships were identified in the α-agarase gene tree. Both *C. echini* A3^T and *Algibacillus agarilyticus* RQJ05^T have two α-agarase genes, which are paralogs, in each
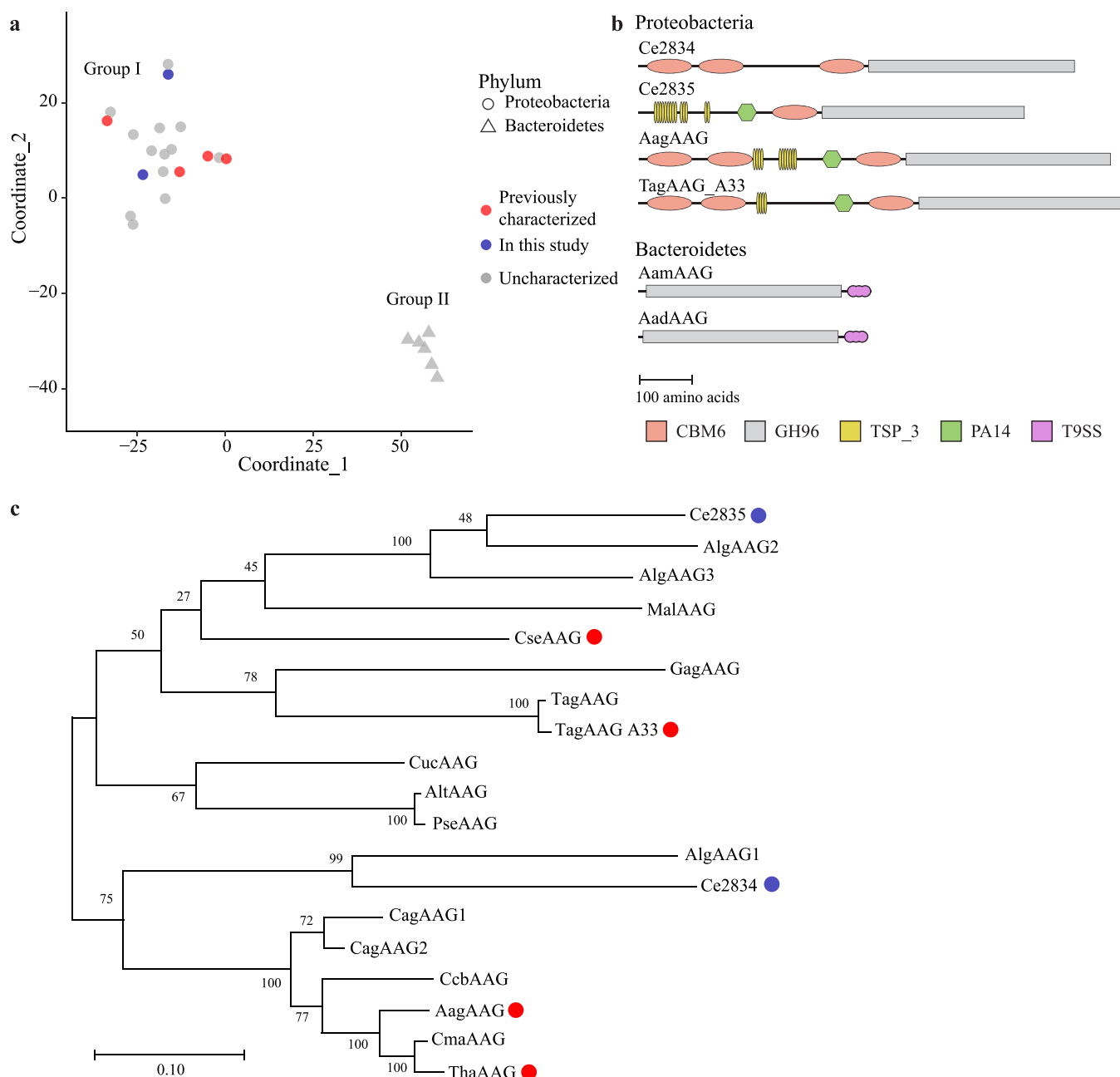
FIG 2 Phylogenetic analysis of GH96 module containing proteins. (a) Multidimensional scaling plot (MDS plot) showing the distance between the sequences containing GH96 catalytic modules in the phyla *Proteobacteria* and *Bacteroidetes*. (b) Modular architecture of representative GH96 catalytic module-containing proteins in the phyla *Proteobacteria* and *Bacteroidetes*. The GH96 catalytic modules, carbohydrate binding module family 6 (CBM6, Pfam module ID 03422) modules, thrombospondin type 3 repeats (TSP3, Pfam module ID 02412), and PA14 modules (Pfam ID 07691) are indicated in the figure. The scale bar shows the length of each protein in the amino acids. (c) Gene tree showing the phylogenetic relationship between GH96 α-agarases Ce2834 and Ce2835 from *C. echini* A3ᵀ and 17 homologs of the phylum *Proteobacteria* retrieved from the NCBI GenBank (as of December 2020). The gene tree was constructed using the maximum-likelihood method, using amino acid sequences. Ce2834 and Ce2835 from *C. echini* A3ᵀ are indicated with blue closed circles. Characterized α-agarases are indicated with red closed circles. The tree was drawn to scale with branch lengths measured in the number of substitutions per site. The symbols and accession numbers of the respective sequences given in parentheses are as follows: Ce2834 (WP_148747643.1), Ce2835 (WP_148747644.1) of *Colwellia echini* A3ᵀ, CagAAG1 (EWH08789.1) of *Catenovulum agarivorans* DS-2, CagAAG2 (WP_040395378.1) of *Catenovulum agarivorans* YM01, CcbAAG (WP_108602099.1) *Catenovulum* sp. CCB-QB4, CmaAAG (WP_048688751.1) of *Catenovulum maritimum* Q1ᵀ, AlgAAG1 (WP_111979920.1), AlgAAG2 (WP_198673760.1), AlgAAG3 (WP_111979921.1) of *Algibacillus agarilyticus* RQJ05ᵀ, CseAAG (WP_143872330.1) of *Catenovulum sediminis* D2ᵀ, CucAAG (WP_076419582.1) of *Colwellia* sp. UCD-KL20, GagAAG (WP_041525041.1) of *Gilvimarinus agarilyticus* M5cᵀ, MalAAG (WP_053979760.1) of *Marinagarivorans algicola* Z1ᵀ, TagAAG (WP_093328038.1), TagAAG_A33 (A1IGV8.1) of *Thalassotalea agarivorans* JAMB A33ᵀ, ThaAAG (AQN80853.1) of *Thalassomonas* sp. LD5, AltAAG (MAD14829.1) of *Alteromonadaceae* bacterium, PseAAG (WP_033185639.1) of *Pseudoalteromonas* sp. PLSV, AagAAG (Q9LAP7.1) of *Alteromonas agarilytica* GJ1B, AamAAG (WP_139195727.1) of *Aquimarina amphilecti* 92Vᵀ, and AadAAG (WP_147404403.1) of *Aquimarina* sp. AD1.
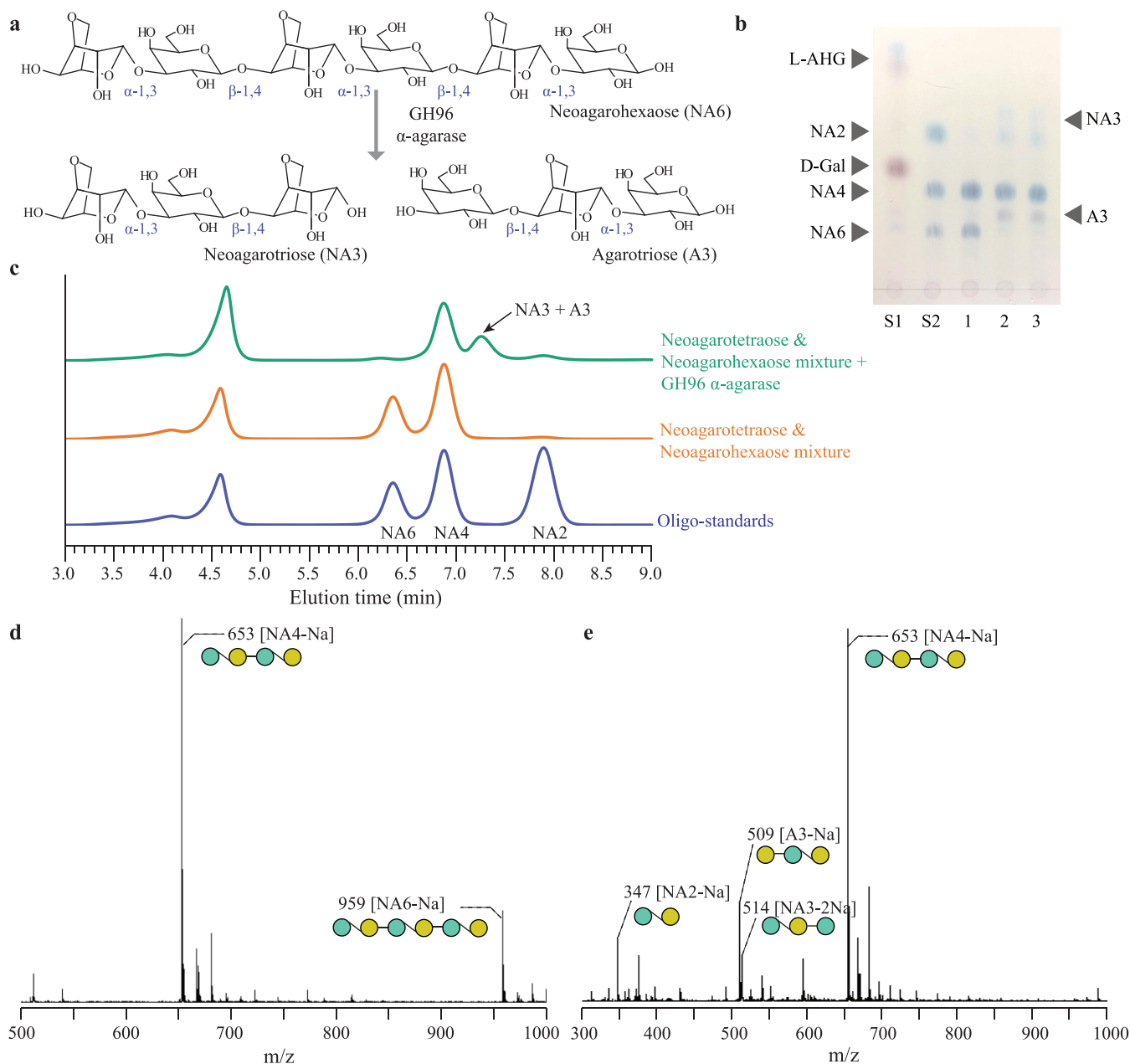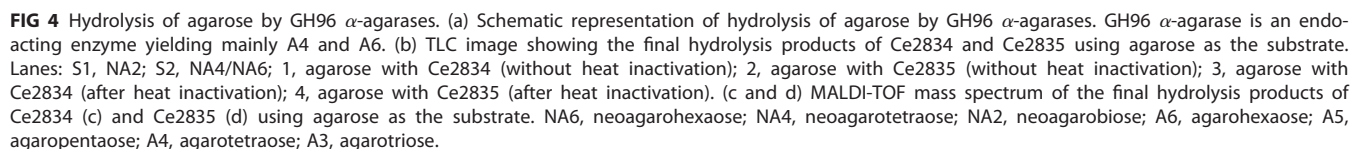
FIG 3 Functional characterization of GH96 α-agarases from *C. echini* A3[T]. (a) Reaction scheme of the GH96 α-agarase hydrolyzing neoagarohexaose (NA6) in forming neoagarotriose (NA3) and agarotriose (A3). (b) TLC showing the hydrolysis products released from NA6 by α-agarase. Lanes: S1, D-Gal and L-AHG; S2, NA2/NA4/NA6; 1, NA4/NA6 mixture with no enzyme; 2, NA4/NA6 mixture with Ce2834; 3, NA4/NA6 mixture with Ce2835. (c) HPLC analysis for the hydrolysis products of NA6 by α-agarase. Oligosaccharide standards (NA2, NA4, and NA6) are shown in blue. (d and e) MALDI-TOF mass spectrum of NA4/NA6 before (d) and after (e) treatment with α-agarase. D-Gal, D-galactose (yellow closed circle); L-AHG, 3,6-anhydro-L-galactose (green closed circle); NA6, neoagarohexaose; NA4, neoagarotetraose; NA3, neoagarotriose; NA2, neoagarobiose; A3, agarotriose.

genome. AlgAAG1 and AlgAAG2 of *Algibacillus agarilyticus* RQJ05[T] are orthologs of Ce2834 and Ce2835, respectively.

The Ce2834 and Ce2835 genes were cloned and the recombinant proteins were expressed in *Escherichia coli*. The endo α-1,3 glycosidase activity of Ce2834 and Ce2835 was confirmed using a mixture of neoagarotetraose (NA4) and neoagarohexaose (NA6) as the substrates. Upon reaction with Ce2834 or Ce2835, NA4 remained intact, whereas NA6 was hydrolyzed to yield two products (Fig. 3a and b). The resulting products were neither D-galactose nor L-AHG, suggesting that the first and third α-1,3 linkages of NA6 were not hydrolyzed. The reaction products resulting from the hydrolysis of the second α-1,3 linkages were identified as

FIG 4 Hydrolysis of agarose by GH96 α-agarases. (a) Schematic representation of hydrolysis of agarose by GH96 α-agarases. GH96 α-agarase is an endo-acting enzyme yielding mainly A4 and A6. (b) TLC image showing the final hydrolysis products of Ce2834 and Ce2835 using agarose as the substrate. Lanes: S1, NA2; S2, NA4/NA6; 1, agarose with Ce2834 (without heat inactivation); 2, agarose with Ce2835 (without heat inactivation); 3, agarose with Ce2834 (after heat inactivation); 4, agarose with Ce2835 (after heat inactivation). (c and d) MALDI-TOF mass spectrum of the final hydrolysis products of Ce2834 (c) and Ce2835 (d) using agarose as the substrate. NA6, neoagarohexaose; NA4, neoagarotetraose; NA2, neoagarobiose; A6, agarohexaose; A5, agaropentaose; A4, agarotetraose; A3, agarotriose.

agarotriose (A3) and neoagarotriose (NA3) (Fig. 3a). High-pressure liquid chromatography (HPLC) analysis supported the observation, showing a decreased NA6 peak and a new peak corresponding to trisaccharide(s) (NA3 or A3) (Fig. 3c). Furthermore, in the matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) analysis, two new peaks appeared at 509 m/z ([A3-Na]) and 514 m/z ([NA3-2Na]), confirming the formation of A3 and NA3 upon hydrolysis of the second α-1,3 linkage in NA6 (Fig. 3d and e). We also observed a peak corresponding to neoagarobiose (NA2) at 347 m/z ([NA2-Na]), which could be the result of the spontaneous release of L-AHG at the reducing end of NA3. A similar spontaneous release of unstable L-AHG moieties from the reducing end of agar-derived oligosaccharides has previously been observed (26).

Ce2834 and Ce2835 showed activity on agarose and primarily yielded agarotetraose (A4) and agarohexaose (A6) (Fig. 4a and b). MALDI-TOF MS of the agarose hydrolysis products from Ce2834 and Ce2835 showed peaks at 653 m/z ([A4-Na]) and 959 m/z ([A6-Na]), which correspond to A4 and A6, respectively (Fig. 4c and d). Product profiles of previously characterized α-agarases support our observation, as α-agarases from *Alteromonas agarilyticus* GJ1B, *T. agarivorans* JAMB-A33^T, and *Thalassomonas* sp. LD5 produced A6 and/or A4 (24–26). Two peaks at 509 m/z ([A3-Na]) and 815 m/z ([A5-Na]), were also observed in both spectra, as previously reported for α-agarases (24, 26). This can be explained by the formation of odd-numbered agarooligosaccharides from

even-numbered ones by the spontaneous hydrolysis of L-AHG at the reducing end (A5 from A6 and A3 from A4), as previously described (Figure 4a) (26).

α-Agarases of *C. echini* A3^T showed optimum activity at 20°C, and the enzymes lost more than 80% of their activities when the temperature exceeded 30°C (Fig. S3). Enzyme activity was not detected at temperatures above 30°C. This is an unusual characteristic compared to previously characterized α-agarases (*Alteromonas agarilyticus* GJ1B, 42.5°C and *Thalassomonas* sp. LD5, 35°C) (24, 26). Recombinant protein constructs, containing only the GH96 catalytic modules, did not show activity on agarose (data not shown). This implies that at least one CBM6 needs to accompany the GH96 catalytic module for the α-agarases to be active.

Molecular characterization of GH86 and GH50 β-agarases. *C. echini* A3^T encodes five β-agarases that belong to two different GH families: GH50 (Ce2839, Ce2840, Ce2842, and Ce2862) and GH86 (Ce2867). Phylogenetic analysis revealed that the GH50 β-agarases of *C. echini* A3^T are close homologs of previously known β-agarases (Table S3, Fig. S4). Similarly, the GH86 β-agarases of *C. echini* A3^T showed high amino acid identity to putative β-agarases in other *Colwellia* species and displayed less amino acid identity with previously characterized GH86 β-agarases (Table S3, Fig. S5). All five β-agarases were cloned and expressed in *E. coli*. Only one of the four GH50 β-agarases (Ce2862) and the GH86 β-agarase (Ce2867) of *C. echini* A3^T were functionally characterized. We were unable to determine the activity of the other GH50 β-agarases as we failed to obtain the soluble and active recombinant enzymes.

β-Agarases of the GH86 family are endo-acting enzymes, whereas GH50 members are exo-acting enzymes (Fig. S6a) (17, 33). Thin-layer chromatography (TLC) analysis of agarose hydrolysates of Ce2867 and Ce2862 showed that NA4 and NA6 are the major products of GH86 β-agarase (Ce2867), while NA2 is the major product of GH50 β-agarase hydrolysis (Ce2862) (Fig. S6b). MALDI-TOF MS analysis of GH86 β-agarase products had two major peaks at 653 *m/z* ([NA4-Na]) and 959 *m/z* ([NA6-Na]) (Fig. S6c), whereas the major peak generated by GH50 β-agarase was at 342 *m/z* ([NA2-Na]) (Fig. S6d). Previously characterized GH86 β-agarases showed similar product profiles as Ce2867 from *C. echini* A3^T, namely, NA4, NA6, and NA8 from Aga86E in *S. degradans* 2-40 (33), NA2 and NA4 from a GH86 β-agarase in a nonmarine agarolytic bacterium *Cellvibrio* sp. OA-2007 (34), and NA6 from a GH86 β-agarase in *Microbulbifer thermotolerans* (35). Exo-acting β-agarases belonging to the GH50 family have been shown to produce NA2 as the major product (17). We confirmed that Ce2862 is an exo-acting β-agarase based on the sequence similarity and product profile of previously characterized GH50 β-agarases. Ce2862 shared 46.9%, 45.8%, and 43.3% amino acid identity with the GH50 β-agarases Ce2839, Ce2840, and Ce2842, respectively. The sequence and biochemical evidence suggest that all four GH50 β-agarases have the same function.

Other agarolytic enzymes and the metabolic fate of L-AHG in *C. echini* A3^T. The AGA PUL of *C. echini* A3^T contains one copy of an agarolytic β-galactosidase (ABG, Ce2828) and an α-neoagarobiose hydrolase (NABH, Ce2875), belonging to GH2 and GH117, respectively (18, 19). The ABG of *C. echini* A3^T (Ce2828) has a 37% amino acid identity to the previously characterized *Vej*ABG of *Vibrio* sp. EJY3. We constructed a phylogenetic tree using close sequence homologs of Ce2828, previously characterized ABGs, and previously characterized β-galactosidase (LacZ). Two distinct clades were observed for ABG and β-galactosidase (LacZ). We found that Ce2828 was grouped with other ABGs (Fig. S7) and Ce2875 has 50% amino acid identity to previously characterized NABH from *S. degradans* 2-40 (Fig. S8) (18). We cloned the genes encoding Ce2828 and Ce2875 and expressed recombinant proteins in *E. coli*. Ce2828 displayed β-galactosidase activity against agarooligosaccharides (AOS) (confirmed with A3) but was inactive on neoagarooligosaccharides (NAOS) (NA2 and NA3) (Fig. S9a and b). This molecular functionality was verified using the previously characterized *Vej*ABG of *Vibrio* sp. EJY3 (Fig. S9c) (19). The activity of Ce2875 in hydrolyzing neoagarobiose to D-galactose and L-AHG was confirmed using the characterized NABH of *S. degradans* 2-40 as a reference (18) (Fig. S10).

The biochemical pathways responsible for metabolizing D-galactose and L-AHG in *C. echini* A3^T were inferred by sequence similarity analysis and supported by the RNA-seq

analysis. D-Galactose is utilized in the Leloir pathway, the predominant route of galactose metabolism (36). The following four enzymes convert L-AHG to 2-keto-3-deoxy-D-gluconate: AHG dehydrogenase (Ce2850), AHGA cycloisomerase (Ce2847), KDG dehydrogenase (Ce2848), and DDG reductase (Ce2849). Subsequent phosphorylation by KDG kinase (Ce2874) and hydrolysis by KDPG aldolase yield pyruvate and glyceraldehyde-3-phosphate, intermediates of the glycolysis (12, 37). The experimentally characterized AHG dehydrogenase and AHGA cycloisomerase from *Vibrio* sp. EJY3 have a 54% sequence identity to Ce2850 and 63% sequence identity to Ce2847, respectively (38). In addition to the sequence similarities, all five genes (Ce2847 to Ce2850 and Ce2874) were located within the PUL and were also among the highly upregulated genes in the RNA-seq profiling. We also observed upregulated sulfatases (Ce2853 and Ce2855) and cytochrome P450 monooxygenases (Ce2880 to Ce2882), involved in the desulfation (39, 40) and demethylation (41) of algal polysaccharides, respectively.

Elucidation of the complete agarolytic pathway in *C. echini* A3. Through the combination of functional annotation, predicted cellular localization (Table S4 and Data S3), and biochemical characterization, we propose a model for the complete agar metabolic system in *C. echini* A3^T (Fig. 5). Cellular localization of the enzymes encoded in the AGA PUL (Table S4 and Data S3) were predicted and used to illustrate the flow of metabolites. Agar depolymerization starts with the membrane-bound, surface-exposed α- and β-agarases. The GH86 β-agarase (Ce2867) depolymerizes agar to NA4 and NA6, followed by further hydrolysis to NA2 by GH50 β-agarases (Ce2839, Ce2840, and Ce2862). Surface-exposed exo-acting GH50 β-agarases also hydrolyze agar into NA2. The α-agarase located on the outer surface acts on two substrates, agar and NAOS, with a degree of polymerization (DP) greater than 6 (e.g., NA6, NA8, and NA10) (Fig. 5, I. Extracellular process). The extracellular α-agarase mainly produces agarooligosaccharides (AOS) with DP4 or DP6 (A4 and A6) using agar as the substrate. Furthermore, the extracellular α-agarase also uses NAOS with DP6 or higher to produce a mixture of odd-numbered agarooligosaccharides (e.g., NA3 and A3 from NA6). The pool of AOS (e.g., A3, A4, and A6) and NAOS (e.g., NA2, NA3, NA4, and NA6), produced by the combined activity of extracellular α-and β-agarases, is transported into the periplasmic space through TBDT (Ce2859 and Ce2863). TBDT are involved in oligosaccharide detection and transport; they are functionally analogous to the SusC/SusD system in *Bacteroidetes* (42–44). Both TBDT are encoded in the AGA PUL and were upregulated during growth on agar.

The agarolytic enzymes located in the periplasm hydrolyze NAOS and AOS to NA2. The GH96 α-agarase (Ce2834), GH50 β-agarase (Ce2842), and ABG (Ce2828) are the key agarolytic enzymes in the periplasmic space (Fig. 5, II. Periplasmic process). NA6 can be further processed by periplasmic GH96 α-agarase, resulting in A3 and NA3, or by periplasmic GH50 β-agarase, yielding NA2. The D-galactose moiety on the nonreducing end of AOS is cleaved by ABG to yield the respective NAOS and D-galactose. NA4 is hydrolyzed by GH50 β-agarase, yielding NA2. By this process, the end products of the periplasmic agarolytic pathway are monomeric sugar (D-galactose), NA2, and NA3. The resultant NA3 is further hydrolyzed by the GH50 β-agarase, resulting in NA2 and L-AHG (Fig. S11). We suggest that the resulting neoagarobiose (NA2) is transported into the cytoplasm by an MFS transporter (45) such as Ce2838, and monosaccharides are transported by ABC transporter permeases (46) (Ce2869 and Ce2870) or the sodium-hexose cotransport protein (Ce2829) (47) in the cytoplasmic membrane. All the oligo- and monosaccharide transporters are located within the PUL and were highly upregulated in RNA-seq profiling, suggesting that they are actively involved in agar metabolism. NA2 is hydrolyzed to D-galactose and L-AHG by NABH (Ce2875) in the cytoplasm. The resulting D-galactose and L-AHG are metabolized as described elsewhere (12, 36, 37).

Core and auxiliary agarolytic pathways define the metabolic versatility of *C. echini* A3^T. Although a few α-agarases have been functionally characterized (24–26), their cellular contribution to the complete hydrolysis of agar into D-galactose and L-AHG by the α-agarolytic pathway has never been described in the literature. Based on the experimental evidence presented here, we propose an extended agarolytic model for marine bacteria that includes the α-agarolytic pathway. The AOS produced by GH96 α-agarases are further hydrolyzed by GH50 β-agarases, ABG, and NABH, as described above. Therefore, the single addition of GH96 α-agarase to the known β-agarolytic pathway (mainly composed
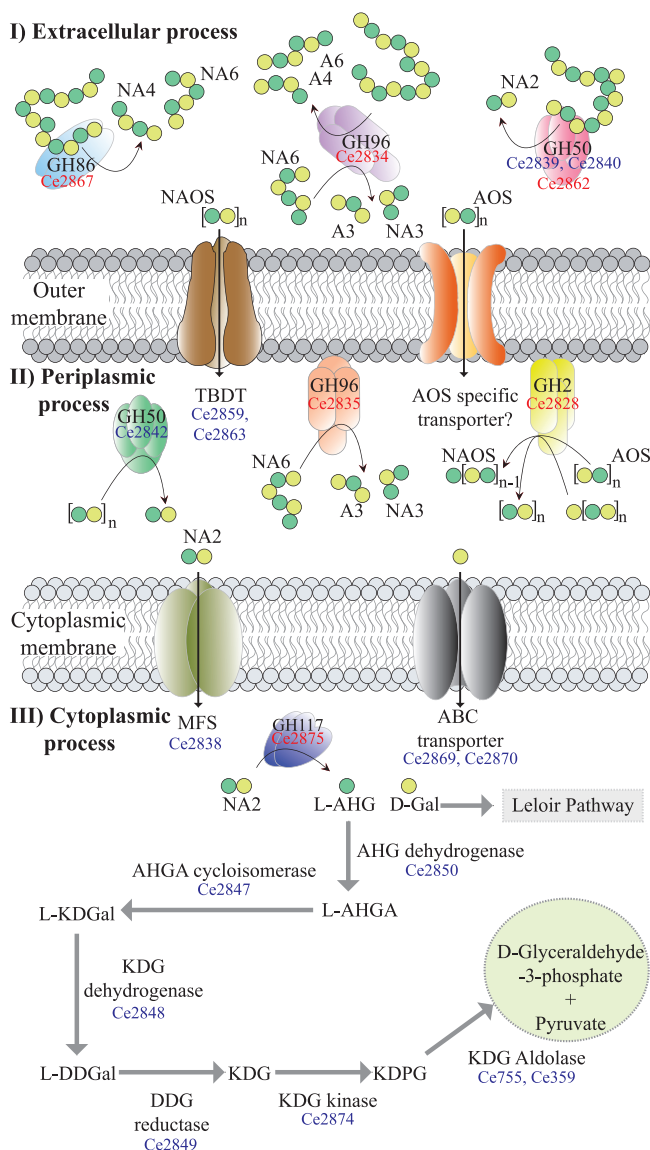
FIG 5 Schematic model of agar metabolism in *C. echini* A3^T. Initial depolymerization occurs at the cell surface by GH86, GH50, and GH96 agarases. Resulting oligosaccharides are funneled into the periplasm by TonB-dependent transporters. Within the periplasm, AOS and NAOS are further hydrolyzed into neoagarobiose by GH2 ABG, GH96, and GH50 agarases. The resulting NA2 and monosaccharides are transported into the cytoplasm by MFS and ABC transporters. NA2 is hydrolyzed to D-galactose and L-AHG by GH117 NABH inside the cytoplasm. L-AHG is metabolized by four enzymes located in the AGA PUL: 3,6-anhydro-α-L-galactose dehydrogenase (AHG dehydrogenase); 3,6-anhydro-α-L-galactonate cycloisomerase (AHGA cycloisomerase); 2-keto-3-deoxy-L-galactonate 5-dehydrogenase (KDG dehydrogenase); and 2,5-diketo-3-deoxy-L-galactonate 5-reductase (DDG reductase). The resulting 2-keto-3-deoxy-D-gluconate is phosphorylated by 2-keto-3-deoxy-D-gluconate kinase (KDG kinase) and hydrolyzed by 2-keto-3-deoxy-D-phosphogluconate aldolase (KDG adolase). D-Glyceraldehyde-3-phosphate and pyruvate enter the glycolysis reaction. D-Galactose is metabolized through the Leloir pathway. Enzymes that are functionally annotated and verified by RNA-seq analysis are indicated in dark blue. Enzymes that are functionally annotated, verified by RNA-seq analysis, and functionally characterized are indicated in red. Gal, D-galactose; L-AHG, 3,6-anhydro-L-galactose; L-AHGA, 3,6-anhydro-L-galactonate; L-KDGal, 2-keto-3-deoxy-L-galactonate; L-DDGal, 2,5-diketo-3-deoxy-L-galactonate; KDG, 2-keto-3-deoxy-D-gluconate; KDPG, 2-keto-3-deoxy-D-phosphogluconate; NA6, neoagarohexaose; NA4, neoagarotetraose; NA3, neoagarotriose; NA2, neoagarobiose; A3, agarotriose; NAOS, neoagarooligosaccharides; AOS, agarooligosaccharides.

of β-agarases, ABG, and NABH) gives rise to a novel auxiliary α-agarolytic pathway in *C. echini* A3^T.

Functional annotation and protein characterization suggest that *C. echini* A3^T uses a complex combination of these agarolytic enzymes. This complex agarolytic system
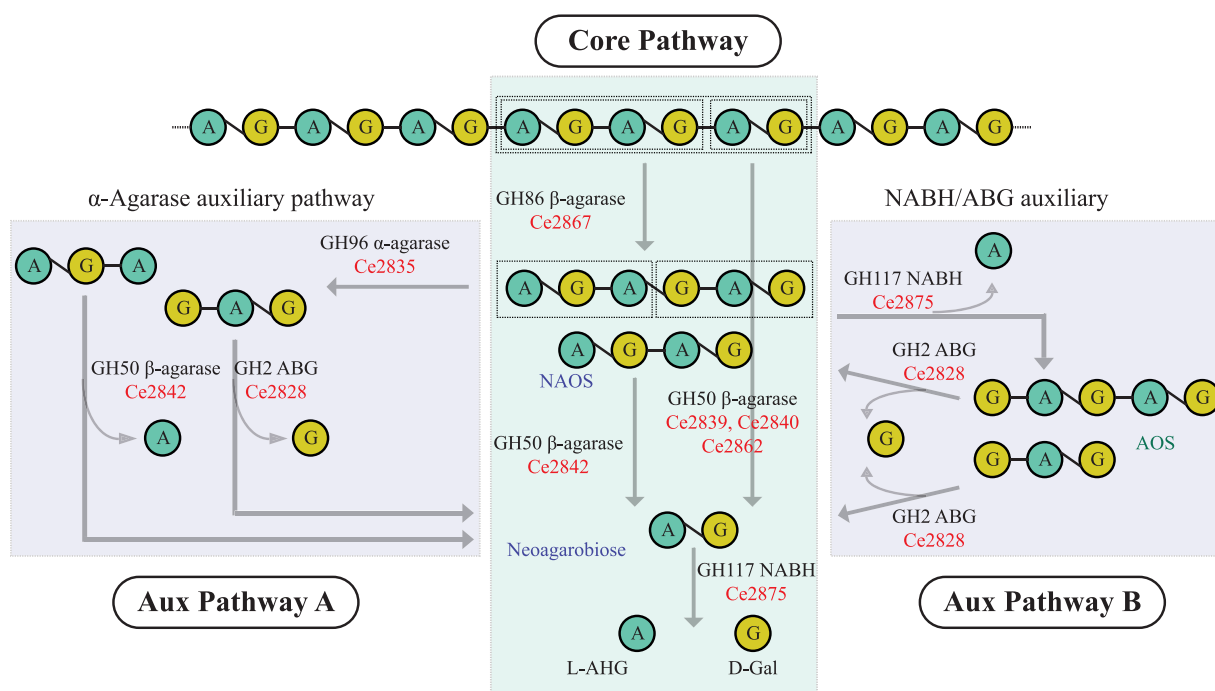
FIG 6 Divergence of the core β-agarolytic pathway and emergence of auxiliary agarolytic pathways in *C. echini* A3^T. Complete hydrolysis of agar into D-gal and L-AHG through the core β-agarolytic pathway (core pathway) in *C. echini* A3^T recruits GH86, GH50 β-agarases, and GH117 NABH. Other agarolytic microorganisms that harbor GH16 and GH118 β-agarases can include these also in the core β-agarolytic pathway. In the ABG/NABH auxiliary pathway (Aux Pathway B), NABH releases the L-AHG at the nonreducing end of even-numbered NAOS and produces odd-numbered AOS. D-Galactose at the nonreducing end of the resulting AOS is released by ABG. In the α-agarase auxiliary pathway (Aux Pathway A), even-numbered NAOS are hydrolyzed to a mixture of odd-numbered AOS and NAOS. These AOS require ABG to release the D-galactose at the nonreducing end and convert them to NAOS. NAOS, neoagarooligosaccharides; AOS, agarooligosaccharides; L-AHG, 3,6-anhydro-L-galactose; D-Gal, D-galactose.

exemplifies the emergence of auxiliary pathways diverging from the core agarolytic pathway in marine agar-degrading bacteria, indicating metabolic versatility (Fig. 6). The core and auxiliary pathways are not entirely compartmentalized, which allows them to share the intermediates (e.g., AOS and NAOS of variable DPs). The core pathway is primarily composed of a combination of various β-agarases and NABH that are indispensable for hydrolyzing agar to D-galactose and L-AHG. The core agarolytic pathway is highly conserved in many microorganisms that feed on agar. In contrast, an auxiliary pathway contains less conserved enzymes that diversify the intermediates of the core pathway and provide alternative routes to obtain monosaccharides.

The core β-agarolytic pathway of *C. echini* A3^T is composed of GH50, GH86 β-agarases, and NABH (Fig. 6, core pathway). The β-galactosidase activity by ABG, along with the NAOS hydrolase activity of NABH, can be regarded as an auxiliary pathway (ABG/NABH auxiliary pathway) for converting NAOS to D-galactose and L-AHG (Fig. 6, Aux Pathway B). NABH releases L-AHG from the nonreducing end of even-numbered NAOS in this auxiliary pathway and produces odd-numbered AOS. Then D-galactose at the nonreducing end of the resulting AOS is released by ABG, producing even-numbered NAOS. This process continues until NAOS is completely hydrolyzed to D-galactose and L-AHG. This system is not unprecedented, as it was reported that *Vibrio* sp. EJY3 can also hydrolyze agar using the same dual agarolytic pathways (core β-agarolytic pathway and ABG/NABH auxiliary pathway) (48). We have elucidated a new auxiliary pathway in *C. echini* A3^T (α-agarase auxiliary pathway) (Fig. 6, Aux Pathway A) in which an α-agarase plays the central role. The α-agarase auxiliary pathway also employs ABG, which links the core β-agarolytic pathway (GH50/86 β-agarases and NABH) to the auxiliary pathway. The addition of a GH96 α-agarase means that NAOS generated by the core pathway provides an influx of intermediates to the α-agarase auxiliary pathway.

In the α-agarase auxiliary pathway, NAOSs are hydrolyzed by α-agarase into a mixture of odd-numbered AOS and NAOS. The AOSs require ABG to release the D-galactose at the nonreducing end and convert them to NAOS, which is further hydrolyzed by GH50 β-agarase and NABH in the core pathway.
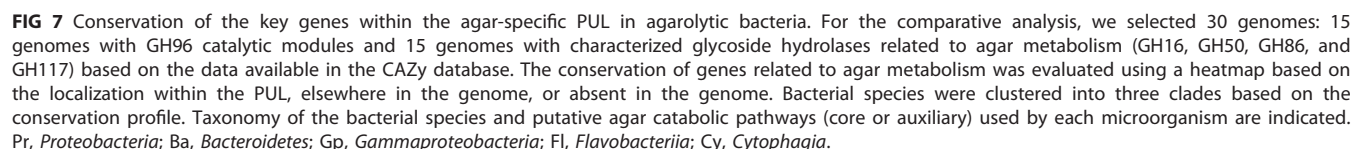
AGA PUL of *C. echini* A3^T encodes both the core and auxiliary agarolytic pathways. The metabolic versatility model that illustrates the sharing of intermediates between core and various auxiliary pathways was confirmed by analysis of the genomes of agarolytic marine microorganisms. We hypothesized that gene gain in AGA PUL is the key event for auxiliary function that drives metabolic versatility. To investigate the metabolic versatility in other marine bacteria, we collected genome sequences of 15 bacteria possessing characterized glycoside hydrolases related to agar metabolism (as in CAZy) and an additional 15 genomes containing GH96 catalytic modules (the list of these bacteria is in Table S5). Once all protein-coding sequences in the 30 genomes were clustered into the orthologous groups based on sequence similarity, we predicted the genomic regions containing potential AGA PULs using the essential genes for agar metabolism (L-AHG metabolic enzymes and NABH) as probes. The boundaries of potential AGA PUL in each genome were manually inspected and ortholog groups related to agar metabolism located within the PUL were annotated (Fig. S12, Table S6, S7, Data S3). The potential AGA PULs of *Bacteroidetes* were cross-checked with the PUL database (PULDB) (49). Our predictions were compatible with the PUL available in the PULDB except for *Aquimarina* sp. BL5, *Aquimarina* sp. AD1, and *Bacteroides plebeius* M12. The genus *Aquimarina* lacked predicted PULs in PULDB and we could not predict the agar-specific PUL in *Bacteroides plebeius* M12.

Based on the composition of conserved genes in the potential PUL, the bacterial genomes were clustered into three clades (Fig. 7). While members in clade 1 and clade 2 belong to *Gammaproteobacteria*, members of clade 3 are from the classes *Flavobacteriia* and *Cytophagia*, belonging to *Bacteroidetes*. The SusC/SusD genes were found exclusively within the phylum *Bacteroidetes*. Transcriptional regulators of the IclR family and the GntR family are highly conserved within the PULs of *Gammaproteobacteria*, whereas *Bacteroidetes* lacked both the GntR family and both transcriptional regulator types. KDG kinase was also not found within the phylum *Bacteroidetes*. Lack of a similar kinase gene in the carrageenan-specific PULs of some *Bacteroidetes* was reported in the literature, and it was suggested that they might use a nonphosphorylation variant of the Entner-Doudoroff pathway (11, 50). We identified two ortholog groups of GH16 β-agarase, representing two different GH16 subfamilies. Sequence analysis using the dbCAN2 meta server (51) showed that one ortholog group represents GH16_15, including AgaC from *Z. galactanivorans* Dsij^T, and the other represents the GH16_16, including Aga16B, from *S. degradans* 2-40. Two ortholog groups were identified for the GH86 β-agarase as well, suggesting two potential, yet undefined, GH86 subfamilies.

β-Agarase and NABH are highly conserved across the genomes studied, emphasizing their role in the core agarolytic pathway. NABH was not found in *Microbulbifer agarylyticus* strain JAMB-A3^T or *Microbulbifer thermotolerans* strain JMAB-A94, although it is plausible that these bacteria have distant homologs that perform the same molecular function. The *Aquimarina* sp. BL5 does not encode any β-agarase, but does code for GH96, ABG, NABH, and L-AHG metabolic enzymes in its genome. The majority of genomes containing β-agarase and/or NABH (23 of 29) possessed ABG. This suggests that microorganisms with ABG diversify the core pathway and evolve the ABG/NABH auxiliary pathway for agar degradation (Fig. 6, Aux Pathway B) (48). We observed the coexistence of ABG and α-agarase in all genomes containing α-agarase (except for *Cellvibrionaceae* bacterium AOL6) suggesting that α-agarase collaborates with ABG to extend the core agarolytic pathway by building a second auxiliary pathway (α-agarase auxiliary pathway) (Fig. 6, Aux Pathway A).

Although known core genes for agar metabolism (β-agarases, NABH, and L-AHG metabolic enzymes) are highly conserved within the PUL, the auxiliary gene composition of the PULs (ABG, α-agarase, etc.) vary among the microorganisms. This evidence strengthens the notion that the core agarolytic system of marine bacteria is regulated together within the agarolytic PUL. While the ABG/NABH auxiliary pathway is linked to

FIG 7 Conservation of the key genes within the agar-specific PUL in agarolytic bacteria. For the comparative analysis, we selected 30 genomes: 15 genomes with GH96 catalytic modules and 15 genomes with characterized glycoside hydrolases related to agar metabolism (GH16, GH50, GH86, and GH117) based on the data available in the CAZy database. The conservation of genes related to agar metabolism was evaluated using a heatmap based on the localization within the PUL, elsewhere in the genome, or absent in the genome. Bacterial species were clustered into three clades based on the conservation profile. Taxonomy of the bacterial species and putative agar catabolic pathways (core or auxiliary) used by each microorganism are indicated. Pr, *Proteobacteria*; Ba, *Bacteroidetes*; Gp, *Gammaproteobacteria*; Fl, *Flavobacteriia*; Cy, *Cytophagia*.

the PUL in most *Bacteroidetes*, the ABG in most *Proteobacteria* was found elsewhere in the genome (i.e., outside the PUL). In most of the α-agarase-encoding microorganisms, α-agarase genes were found outside the PUL. However, *C. echini* A3^T has a highly inclusive PUL in which all genes necessary for core and auxiliary agarolytic pathways, including ABG and α-agarase, are enclosed. We did not find similar PUL organization in any other genomes studied here. The PUL organization of *C. echini* A3^T indicates that these genes are collectively regulated as a functional unit handling agar hydrolysis, transport, and metabolism in a concerted manner, which was also supported by RNA-seq analysis (Fig. 1b).

DISCUSSION

The α-agarases have rarely been found in the known PULs of agarolytic bacteria (49). Thus, we aimed to characterize the AGA PUL of *C. echini* A3^T, which encodes two GH96

α-agarases. The genomic, transcriptomic, and biochemical analyses presented here support the observation that the agarolytic system of *C. echini* A3^T is arranged and regulated similar to PULs in *Bacteroidetes*. As such, this study supports the notion that PUL-like gene clusters are not limited to the phylum *Bacteroidetes*, but are also found in *Gammaproteobacteria* known to lack the canonical SusC/SusD genes (11, 12). Comparative genomic analysis of 30 marine agarolytic bacteria in this study revealed that potential AGA PULs are widely distributed in marine agarolytic bacteria, which was validated with the experimentally characterized PULs in the database (49). The sizes and composition of these potential AGA PULs varied from one species to another (sizes of 8 to ~110 kb representing 6 to ~70 genes). While the core agarolytic pathway, exemplified by β-agarase, NABH, and L-AHG metabolic enzymes, is conserved in marine bacteria, the addition of genes to this core agarolytic repertoire has allowed auxiliary pathways to evolve (i.e., α-agarases in this study and ABG in reference 48). Such evolutionary events eventually contribute to microbial adaptation and colonization into diverse environmental niches. In *C. echini* A3^T, the diversification of the AGA PUL is highly evident, as it contains the genes encoding both core and auxiliary functions within the same gene cluster.

Genomic analysis of marine bacteria showed that α-agarases are rare in natural environments. They are distributed almost exclusively among closely related orders of *Gammaproteobacteria* (*Alteromonadales* and *Cellvibrionales*), suggesting that α-agarases have evolved more recently than β-agarases. The limited distribution of α-agarases in nature also suggests that α-agarases may confer a niche-specific function rather than supporting general agar metabolism. *Colwellia* species are not frequently found on the surface of macroalgae but they are common in cold marine environments and found as symbionts of marine organisms (52). *C. echini* A3^T was originally isolated from the sea urchin *Strongylocentrotus droebachiensis* that relies on an algal diet (28). A mutualistic relationship has been suggested for some endosymbiotic bacteria, isolated from marine herbivores, that produce enzymes to degrade marine algal polysaccharides (53, 54). Although the composition and structure of the microbiome of such herbivores are poorly understood, the nutrient-niche hypothesis predicts that nutrient availability defines the ecological niches within the digestive system (55). Diversification of the agarolytic gene repertoire may allow *C. echini* A3^T to effectively utilize the algal polysaccharides within the sea urchin intestine. Bacteria that can utilize the limiting nutrients can successfully colonize and adapt into these niches, giving them an advantage over other members of the gut microbiome (55–57). We speculate that the alternative degradative ability of α-agarases might facilitate the establishment of *C. echini* A3^T in a specific niche within the sea urchin intestine by selectively utilizing certain agar-derived oligosaccharides.

MATERIALS AND METHODS

Genome sequencing, annotation, and phylogenetic analysis. The draft genome sequence of isolate *C. echini* A3^T is available in the NCBI Reference Sequence database under accession number GCF _002843355.2. Automatic annotation of the genome sequence was performed using Rapid Annotation on the Subsystem Technology server (RAST) (58). Identification of CAZymes and sulfatases was carried out using hidden Markov model searches against the local version of dbCAN database, as described previously (29). GH catalytic modules and noncatalytic CBM6 modules of CAZymes were predicted using the dbCAN2 server (51) and NCBI conserved domain search tool. (www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Signal peptides were predicted using the SignalP 5.0 (59) (http://www.cbs.dtu.dk/services/SignalP/) and LipoP 1.0 (60) (www.cbs.dtu.dk/services/LipoP) servers. Predictions of subcellular localization of proteins were performed using pSORTb v.3.0 (61) (www.psort.org/psortb/) and CELLO v.2.5 (62) (cello.life.nctu.edu.tw). Phylogenetic analysis of α- and β-agarases was done using GH96, GH50, and GH86 sequences available from the NCBI GenBank database (sequences were retrieved in December 2020). Multiple sequence alignments of amino acid sequences were performed using the ClustalOmega tool (https://www.ebi.ac.uk/Tools/msa/clustalo/). The pairwise distance matrix of the amino acid sequences of GH96 catalytic module containing proteins was computed using the Protdist program (Version 3.66) PHYLIP package (63). Phylogenetic analyses were conducted using MEGAX (64) with the maximum-likelihood method with 100 bootstrap iterations.

RNA-seq expression profiling for different carbon sources. For routine culture maintenance, *C. echini* A3^T was grown in Difco Marine Broth 2216 (BD Biosciences, San Jose, CA) at 20°C, in a shaking incubator at 180 rpm. For the transcriptomic analysis, an overnight culture of *C. echini* A3^T in marine broth was harvested and washed three times using phosphate-buffered saline solution to remove any residual marine broth. Washed cells were resuspended and diluted to an optical density at 600 nm (OD$_{600}$) of

~0.1 in marine minimal broth (MMB) containing 2.3% (wt/vol) aquarium sea salt mix (Instant Ocean Sea Salts; Aquarium Systems, Mentor, OH, USA), 0.1% (wt/vol) yeast extract (BD Biosciences, San Jose, CA), 0.05% (wt/vol) NH$_4$Cl (Tokyo Chemical Industries, Japan), and 10 mM Tris-HCl buffer (pH 7.4) (Duchefa Biochemie, Haarlem, Netherlands). Diluted cells were inoculated in triplicate into MMB supplemented with 0.2% D-glucose (Deajung, South Korea) and incubated at 20°C, 180 rpm shaking incubator until reaching the stationary phase. Cells were harvested, washed as described above, and diluted up to OD$_{600}$ ~0.1 in MMB. These cells were used to inoculate 10 ml of MMB supplemented with 0.2% D-glucose (Deajung, South Korea) or 0.2% agar (BD Biosciences, San Jose, CA) as the carbon source. Cells were incubated at 20°C in a 180 rpm shaking incubator until the OD$_{600}$ reached 0.8. Bacterial cultures were mixed well and aliquots of 500 μl of cells were harvested by centrifugation at 4°C. Cell pellets were immediately frozen in liquid nitrogen and stored at −80°C.

Cell pellets were resuspended in 1 ml of TRIzol reagent (Invitrogen, Carlsbad, CA, USA) and incubated for 2 min at room temperature for lysing. Cell lysate was mixed with 200 μl of chloroform and the aqueous phase was separated by centrifugation at 12,500 × g for 5 min. Isopropanol (0.5 volumes) (Sigma, St. Louis, MO, USA) was added to the separated aqueous phase and loaded onto an RNeasy mini column (Qiagen, Germany). Columns were washed with 70% ethanol (Sigma, St. Louis, MO, USA) and the RNA was eluted with RNase-free water. The DNase treatment was performed using RNase-free DNase I (Thermo Fisher Scientific, MA, USA) and the RNA was purified using the RNeasy minikit (Qiagen, Germany). The RNA concentration was determined using a Qubit RNA BR assay kit (Invitrogen, Carlsbad, CA, USA). The quality of RNA was determined using the Qsep1 Bio-fragment analyzer (Bioptic Inc, Taiwan) equipped with an RNA cartridge.

RNA-seq library preparation was performed as follows. mRNA enrichment was performed using the Ribo-Zero rRNA removal kit for bacteria (Illumina, San Diego, CA, USA). rRNA-depleted samples were purified using Agencourt AMPure XP beads (Beckman Coulter, Brea, CA, USA). The RNA-seq libraries were prepared using NEBNext Ultra RNA library prep kit for Illumina (New England BioLabs, Ipswich, MA, USA) according to the manufacturer's instructions. Sequencing was performed on the Illumina MiSeq platform (Illumina, San Diego, CA, USA) using an Illumina MiSeq reagent kit V3 (300 bp paired ends).

Quality control and trimming of Illumina sequence reads were performed with TrimGalore ver. 0.5.0 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). The Illumina universal adapter sequences and the index sequences were trimmed. Reads shorter than 40 bp were discarded and the bases with Q < 20 were trimmed from the 3′ and 5′ ends of the reads. The reads were mapped to the C. echini A3^T genome (GCF _002843355.2) by Bowtie2 software (65). SAMtools (66) was used to convert the alignment files into BAM files. The number of reads mapped to the predicted CDS was analyzed using the Bioconductor package (67). Differential gene expression analysis was performed with the edgeR package (68) in Bioconductor. Genes with a P value of <0.05 and log fold change (logFC) of >2 were considered differentially expressed.

Cloning and expression of enzymes. Genomic DNA of C. echini A3^T was isolated using Exgene Cell SV genomic DNA extraction kit (GeneAll, South Korea). Genes encoding two putative α-agarases (Ce2834 and Ce2835), a GH86 β-agarase (Ce2867), four GH50 β-agarases (Ce2839, Ce2840, Ce2842, and Ce2862), a GH2 agarooligosacchrolytic β-galactosidase (Ce2828), and a GH117 neoagarobiose hydrolase (Ce2875) were cloned in this study. If necessary, secretory signal peptides were removed before cloning. Ce2834 was cloned as four different constructs: (i) the complete Ce2834 gene; (ii) Ce2834 without the signal peptide; (iii) the GH96 catalytic module with the immediately adjacent CBM6 module; and (iv) the GH96 catalytic module without CBM6 modules. Ce2835 was cloned as three different constructs: (i) the complete Ce2835 gene; (ii) the GH96 catalytic module with the CBM6 module; and (iii) the GH96 catalytic module without CBM6 modules. Ce2828, Ce2834, Ce2835, and Ce2875 were cloned into modified pET21a vectors (pB3-His6, pB4-MBP-His6, and pB6-TRX-His6), which were kindly provided by the Structural Genomics Center, UC Berkeley, CA, USA. Ce2839, Ce2840, Ce2842, and Ce2862 were cloned into pET9a-USER vectors as previously described (37). The primer sequences used to amplify the target genes and their respective vectors are listed in Table 1. Bacterial strains used in this study are listed in Table 2.

Recombinant proteins were expressed in *Escherichia coli* strain BL21(DE3). The cells were grown in lysogeny broth (LB) medium (BD Biosciences, San Jose, CA) containing 100 μg/ml ampicillin (Duchefa Biochemie, Haarlem, Netherlands) until the optical density at 600 nm reached 0.5 to 1.0. Overexpression of recombinant proteins was induced by adding 1 M isopropyl-β-D-thiogalactopyranoside (IPTG) (Duchefa Biochemie, Haarlem, Netherlands) to a final concentration of 1 mM. The cells were incubated at 18°C overnight in a 180-rpm shaking incubator.

Preparation of soluble enzyme extracts. Cells were harvested by centrifugation at 5,000 × g for 20 min at 4°C. The cell pellets were resuspended in a lysis buffer containing 20 mM Tris-Cl (pH 8.0), 200 mM NaCl, 5 mM β-mercaptoethanol, and 5% glycerol. The cells were then disrupted by sonication. The soluble and insoluble fractions were separated by centrifugation at 15,400 × g for 20 min at 4°C and the abundance of each recombinant protein in the soluble and insoluble fractions was assessed by SDS-PAGE. The amount of total protein in the soluble extract was determined by Bradford protein assay (Bio-Rad, Hercules, CA, USA) using bovine serum albumin (TaKaRa, CA, USA) as the standard. The soluble enzyme extract was stored at −20°C for activity assay studies.

Enzyme activity assays. Neoagarooligosaccharide substrates were prepared in-house using recombinant GH16 and GH50 β-agarases from S. degradans 2-40, as described earlier (17, 69). Agarose (BD Biosciences, San Jose, CA) was hydrolyzed with β-agarases to generate neoagarobiose, neoagarotetraose, and neoagarohexaose. The hydrolysis products were purified and used in subsequent enzymatic assays. The activity of α-agarases was determined using 1% (wt/vol) agarose (20 mM Tris-Cl, pH 7.5) and a mixture of neoagarotetraose and neoagarohexaose in 20 mM Tris-Cl (pH 7.5). All reaction mixtures were incubated at 20°C overnight. The activities of all C. echini A3^T β-agarases were determined using 1% (wt/vol) agar (20 mM Tris-Cl, pH 7.5) and the reaction was carried out at 30°C overnight. The activity of neoagarobiose hydrolase (NABH) was determined using neoagarobiose

TABLE 1 Primer sequences used for recombinant protein expression in this study

Locus tag	Primer name	Sequence (5′–3′)	Vector(s)
Ce2828	ABG2828_pB_F	GGCGGTGGTGGCGGCATGAGTAAATTAAACGATAATCAGCAGGTA	pB3-His6; pB4-MBP-His6; pB6-TRX-His6
	ABG2828_pB_R	GTTCTTCTCCTTTGCGCCCCTATTACTTTACTTTTATGTAAATGATTTGATCT	
Ce2834	aAg2834_full_pB_F	GGCGGTGGTGGCGGCATGTATAATACAAAACAAACACTAATCAGTGC	pB3-His6; pB4-MBP-His6; pB6-TRX-His6
	aAg2834_WOS_pB_F	GGCGGTGGTGGCGGCATGTATGCAGATACGAGCACAATTCAAG	
	aAg2834_trunc_pB_F	GGCGGTGGTGGCGGCATGTCGAAAGGAGTAGAAAATGGTGATG	
	aAg2834_GH96_pB_F	GGCGGTGGTGGCGGCATGGATCATTATGTAGCTGAAGTTCAAGG	
	aAg2834_pB_R	GTTCTTCTCCTTTGCGCCCCTAATGAGCTAATTCTAATATACCCCACA	
Ce2835	aAg2835_Full_pB_F	GGCGGTGGTGGCGGCATGAAAACATCAAAAATACTACTCTTTTCAAC	pB3-His6; pB4-MBP-His6; pB6-TRX-His6
	aAg2835_Trunc_pB_F	GGCGGTGGTGGCGGCATGTCAGAAGGTGAGGAAACAGG	
	aAg2835_GH96_pB_F	GGCGGTGGTGGCGGCATGGATCATTTTGTAGCCGAAATTGAAG	
	aAg2835_pB_R	GTTCTTCTCCTTTGCGCCCCTATTAGTGACCTAATTCAAGAATACC	
Ce2839	bAg2839_USER_F	GGCTTAAU TTGACAGACAACAAAGATGC	pET9a-USER-His
	bAg2839_USER_R	GGTTTAAU ACCATATCGATTTTGATACATAGTAG	
Ce2840	bAg2840_USER_F	GGCTTAAU ACTGATAATTCAACAGCTATAGTCG	pET9a-USER-His
	bAg2840_USER_R	GGTTTAAU ATCGTTAGTAAAACGTCTGTTATAGA	
Ce2842	bAg2842_USER_F	GGCTTAAU AATACACCAGCAGAGTTAAAGC	pET9a-USER-His
	bAg2842_USER_R	GGTTTAAU ATGTGCATTAGGTAGCATTTC	
Ce2862	bAg2862_USER_F	GGCTTAAU CTTAAATCCACACTTGATGCTA	pET9a-USER-His
	bAg2862_USER_R	GGTTTAAU ATTAAACCGACGTTCATACATC	
Ce2867	bAg2867_pB_F	GGCGGTGGTGGCGGCATGGATGATGAACCATTAGCGCCAAC	pB3-His6; pB4-MBP-His6; pB6-TRX-His6
	bAg2867_pB_R	GTTCTTCTCCTTTGCGCCCCTATTATTTACTTCTTGATAACGCACGACT	
Ce2875	NABH2875_pB_F	GGCGGTGGTGGCGGCATGATGAATTTATCAAATAAAAAAATTAAGTTTAGCAAG	pB3-His6; pB4-MBP-His6; pB6-TRX-His6
	NABH2875_pB_R	GTTCTTCTCCTTTGCGCCCCTACTAATTAGAGTTTTGGAAAGTACCAG	

in 20 mM Tris-Cl (pH 7.5). The NABH enzymatic reaction was carried out at 25°C overnight. A mixture of neoagarotetraose, neoagarotriose, and agarotriose (20 mM Tris-Cl, pH 7.5) was used to determine the specificity of GH2 agarooligosaccharolytic β-galactosidase. The enzymatic reaction was carried out at 25°C overnight.

Qualitative analysis of reaction products by TLC, HPLC, and MALDI-TOF MS. Thin-layer chromatography (TLC) was performed to identify the reaction products of each enzymatic reaction. The enzymatic reaction mixture was centrifuged and 1 μl of supernatant was spotted on a TLC silica gel 60 plate (Merck, Darmstadt, Germany). TLC plates were developed using *n*-butanol:ethanol:water (3:1:1, vol/vol/vol) as the mobile phase. The developed TLC plate was dried, and separated sugars were visualized using 10% (vol/vol) sulfuric acid and 0.2% (wt/vol) naphthoresorcinol (Sigma-Aldrich, St. Louis, MO, USA) in ethanol, followed by heating the TLC plate at 90°C for 2 min.

High-pressure liquid chromatography (HPLC) analysis was performed as follows. Reaction products obtained by the hydrolysis of a neoagarohexaose/neoagarotetraose mixture were analyzed using a Waters 1525 binary HPLC system (Waters) equipped with a gel permeation column (KS-802; Shodex) and a refractive index detector. Analysis was conducted at 50°C using distilled water as the mobile phase at a flow rate of 1.0 ml/min. Molecular size markers of 324 Da, 630 Da, and 936 Da were used to distinguish the degree of polymerization (DP) of products.

TABLE 2 Bacterial strains and vectors used in this study

Strain or vector	Description	Reference or source
Strains		
E. coli DH5α	F⁻ *endA1 supE44 thi-1 recA1 relA1 gyrA96 deoR nupG* Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*) U169 *hsdR17* (r_K^-, m_K^+) λ^-	Invitrogen, Carlsbad, CA, USA
E. coli BL21(DE3)	F⁻ *ompT gal dcm lon hsdS*$_B$ (r_B^- m_B^-) λ (DE3*[lacI lacUV5T7p07 ind1 sam7 nin5]*) [*malB*⁺] $_{K-12}$(λ^S)	Invitrogen, Carlsbad, CA, USA
Colwellia echini A3^T	Gram-negative, rod-shaped, and facultatively anaerobic bacterium that uses agar as sole carbon source.	(28)
Vectors		
pB3-His6	Modified pET21a vector; a LIC sequence with an N-terminus TEV cleavage site and His tag is inserted.	Structural Genomics Center, UC Berkley
pB4-MBP-His6	Modified pET21a vector; a LIC sequence with an N-terminus TEV cleavage site, His tag, and maltose-binding protein fusion tag is inserted.	Structural Genomics Center, UC Berkley
pB6-TRX-His6	Modified pET21a vector; a LIC sequence with an N-terminus TEV cleavage site, His tag, and thioredoxin fusion tag is inserted.	Structural Genomics Center, UC Berkley
pET9a-USER-His	Modified pET21a vector for USER cloning.	(12)

Matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) analysis of the reaction products was carried out as follows. An aliquot of 2 μl of the sample was dropped on the target plate and mixed with 2,5-dihydroxybenzoic acid (DHB) as the matrix. DHB was dissolved in 50:50 (vol/vol) acetonitrile and water containing 0.5% trifluoroacetic acid (TFA) at a concentration of 10 mg/ml. The MS spectra were obtained using a Bruker UltrafleXtreme matrix-assisted laser desorption ionization (MALDI) MS instrument (Bruker, Bremen, Germany) equipped with a SmartBeam II laser.

Bioinformatic analysis of PUL. We searched the GenBank database for bacteria that contain GH96 catalytic modules, based on sequence similarity to Ce2834 and Ce2835. We retrieved 15 genomic sequences that contain GH96 catalytic modules. Genome sequences of *Alteromonas agarlyticus* GJ1B and *Thalassomonas* sp. LD5, which were used to characterize α-agarase, are not available yet. Further, we selected 15 microorganisms that possess characterized glycoside hydrolases related to agar metabolism (GH16, GH50, GH86, and GH117) based on the data available in the CAZyme database. Collectively, 30 genome sequences were used for comparative genome analysis (Table S5) and downloaded from the NCBI RefSeq database. The coding sequences of these 30 genomes were divided into ortholog groups based on sequence similarity using OrthoFinder (70). Using the NABH and L-AHG metabolic enzymes (AHG dehydrogenase, AHGA cycloisomerase, KDG dehydrogenase, and DDG reductase) as queries, the genomic regions that possess the agar-specific PUL were searched. The boundaries of the PUL regions were manually expanded by 5 kb until no additional coding sequences related to agar metabolism (GH16, GH50, GH86, GH96, GH117, GH2, TBDR, or SusC/SusD) were found. Finally, the boundaries of the putative PUL were manually refined. All the ortholog groups that were found within the putative PULs were collected and ortholog groups that were functionally related to agar metabolism (empirical or transcriptomic evidence), or else highly conserved within the PUL, were sorted. We filtered 26 ortholog groups that fall under the criteria above (Table S6). Results were gathered in a matrix to indicate the presence of each ortholog group. Values were assigned as an ortholog group within PUL (value 2), ortholog group elsewhere in the genome (value 1), or ortholog group absent in the genome (value 0). From this matrix, a heatmap and a hierarchical classification of the organisms were made using *pheatmap* R package (71). For hierarchical classification, "complete" algorithm was used with Euclidean distances. Subfamily classification of the GH16 β-agarase was performed using the dbCAN2 meta server (HMMdb v9) (51).

Data availability. The draft genome sequence of *C. echini* A3^T is available in the NCBI Reference Sequence database under the accession number GCF_002843355.2. RNA-seq data are available in the NCBI Sequence Read Archive under accession numbers SRR13994254 to SRR13994259.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.
SUPPLEMENTAL FILE 1, PDF file, 6 MB.
SUPPLEMENTAL FILE 2, XLSX file, 0.1 MB.

ACKNOWLEDGMENTS

REFERENCES

1. Smith S. 1981. Marine macrophytes as a global carbon sink. Science 211:838–840. https://doi.org/10.1126/science.211.4484.838.
2. Martin M, Portetelle D, Michel G, Vandenbol M. 2014. Microorganisms living on macroalgae: diversity, interactions, and biotechnological applications. Appl Microbiol Biotechnol 98:2917–2935. https://doi.org/10.1007/s00253-014-5557-2.
3. Popper ZA, Michel G, Hervé C, Domozych DS, Willats WG, Tuohy MG, Kloareg B, Stengel DB. 2011. Evolution and diversity of plant cell walls: from algae to flowering plants. Annu Rev Plant Biol 62:567–590. https://doi.org/10.1146/annurev-arplant-042110-103809.
4. Kloareg B, Quatrano R. 1988. Structure of the cell walls of marine algae and ecophysiological functions of the matrix polysaccharides. Oceanogr Mar Biol 26:259–315.
5. Usov A. 1992. Sulfated polysaccharides of the red seaweeds. Food Hydrocoll 6:9–23. https://doi.org/10.1016/S0268-005X(09)80055-6.
6. Hehemann J-H, Boraston AB, Czjzek M. 2014. A sweet new wave: structures and mechanisms of enzymes that digest polysaccharides from marine algae. Curr Opin Struct Biol 28:77–86. https://doi.org/10.1016/j.sbi.2014.07.009.
7. Bjursell MK, Martens EC, Gordon JI. 2006. Functional genomic and metabolic studies of the adaptations of a prominent adult human gut

symbiont, *Bacteroides thetaiotaomicron*, to the suckling period. J Biol Chem 281:36269–36279. https://doi.org/10.1074/jbc.M606509200.

8. Martens EC, Koropatkin NM, Smith TJ, Gordon JI. 2009. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. J Biol Chem 284:24673–24677. https://doi.org/10.1074/jbc.R109.022848.

9. Hemsworth GR, Déjean G, Davies GJ, Brumer H. 2016. Learning from microbial strategies for polysaccharide degradation. Biochem Soc Trans 44:94–108. https://doi.org/10.1042/BST20150180.

10. Grondin JM, Tamura K, Déjean G, Abbott DW, Brumer H. 2017. Polysaccharide utilization loci: fueling microbial communities. J Bacteriol 199: e00860-16. https://doi.org/10.1128/JB.00860-16.

11. Ficko-Blean E, Préchoux A, Thomas F, Rochat T, Larocque R, Zhu Y, Stam M, Génicot S, Jam M, Calteau A, Viart B, Ropartz D, Pérez-Pascual D, Correc G, Matard-Mann M, Stubbs KA, Rogniaux H, Jeudy A, Barbeyron T, Médigue C, Czjzek M, Vallenet D, McBride MJ, Duchaud E, Michel G. 2017. Carrageenan catabolism is encoded by a complex regulon in marine heterotrophic bacteria. Nat Commun 8:1685. https://doi.org/10.1038/s41467-017-01832-6.

12. Schultz-Johansen M, Bech PK, Hennessy RC, Glaring MA, Barbeyron T, Czjzek M, Stougaard P. 2018. A novel enzyme portfolio for red algal polysaccharide degradation in the marine bacterium *Paraglaciecola hydrolytica* S66ᵀ encoded in a sizeable polysaccharide utilization locus. Front Microbiol 9:839. https://doi.org/10.3389/fmicb.2018.00839.

13. Cuskin F, Lowe EC, Temple MJ, Zhu Y, Cameron E, Pudlo NA, Porter NT, Urs K, Thompson AJ, Cartmell A, Rogowski A, Hamilton BS, Chen R, Tolbert TJ, Piens K, Bracke D, Vervecken W, Hakki Z, Speciale G, Munoz-Munoz JL, Day A, Peña MJ, McLean R, Suits MD, Boraston AB, Atherly T, Ziemer CJ, Williams SJ, Davies GJ, Abbott DW, Martens EC, Gilbert HJ. 2015. Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. Nature 517:165–169. https://doi.org/10.1038/nature13995.

14. Larsbrink J, Rogers TE, Hemsworth GR, McKee LS, Tauzin AS, Spadiut O, Klinter S, Pudlo NA, Urs K, Koropatkin NM, Creagh AL, Haynes CA, Kelly AG, Cederholm SN, Davies GJ, Martens EC, Brumer H. 2014. A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. Nature 506:498–502. https://doi.org/10.1038/nature12907.

15. Araki C. 1956. Structure of the agarose constituent of agar-agar. BCSJ 29:543–544. https://doi.org/10.1246/bcsj.29.543.

16. Ekborg NA, Taylor LE, Longmire AG, Henrissat B, Weiner RM, Hutcheson SW. 2006. Genomic and proteomic analyses of the agarolytic system expressed by *Saccharophagus degradans* 2-40. Appl Environ Microbiol 72:3396–3405. https://doi.org/10.1128/AEM.72.5.3396-3405.2006.

17. Kim HT, Lee S, Lee D, Kim HS, Bang WG, Kim KH, Choi IG. 2010. Overexpression and molecular characterization of Aga50D from *Saccharophagus degradans* 2-40: an exo-type beta-agarase producing neoagarobiose. Appl Microbiol Biotechnol 86:227–234. https://doi.org/10.1007/s00253-009-2256-5.

18. Ha SC, Lee S, Lee J, Kim HT, Ko HJ, Kim KH, Choi IG. 2011. Crystal structure of a key enzyme in the agarolytic pathway, alpha-neoagarobiose hydrolase from *Saccharophagus degradans* 2–40. Biochem Biophys Res Commun 412:238–244. https://doi.org/10.1016/j.bbrc.2011.07.073.

19. Lee CH, Kim HT, Yun EJ, Lee AR, Kim SR, Kim J-H, Choi I-G, Kim KH. 2014. A novel agarolytic *β*-galactosidase acts on agarooligosaccharides for complete hydrolysis of agarose into monomers. Appl Environ Microbiol 80:5965–5973. https://doi.org/10.1128/AEM.01577-14.

20. Hehemann JH, Correc G, Thomas F, Bernard T, Barbeyron T, Jam M, Helbert W, Michel G, Czjzek M. 2012. Biochemical and structural characterization of the complex agarolytic enzyme system from the marine bacterium *Zobellia galactanivorans*. J Biol Chem 287:30571–30584. https://doi.org/10.1074/jbc.M112.377184.

21. Roh H, Yun EJ, Lee S, Ko HJ, Kim S, Kim BY, Song H, Lim KI, Kim KH, Choi IG. 2012. Genome sequence of *Vibrio* sp. strain EJY3, an agarolytic marine bacterium metabolizing 3,6-anhydro-L-galactose as a sole carbon source. J Bacteriol 194:2773–2774. https://doi.org/10.1128/JB.00303-12.

22. Vera J, Alvarez R, Murano E, Slebe J, Leon O. 1998. Identification of a marine agarolytic *Pseudoalteromonas* isolate and characterization of its extracellular agarase. Appl Environ Microbiol 64:4378–4383. https://doi.org/10.1128/AEM.64.11.4378-4383.1998.

23. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–D495. https://doi.org/10.1093/nar/gkt1178.

24. Potin P, Richard C, Rochas C, Kloareg B. 1993. Purification and characterization of the *α*-agarase from *Alteromonas agarlyticus* (Cataldi) comb. nov., strain GJ1B. Eur J Biochem 214:599–607. https://doi.org/10.1111/j.1432-1033.1993.tb17959.x.

25. Ohta Y, Hatada Y, Miyazaki M, Nogi Y, Ito S, Horikoshi K. 2005. Purification and characterization of a novel alpha-agarase from a *Thalassomonas* sp. Curr Microbiol 50:212–216. https://doi.org/10.1007/s00284-004-4435-z.

26. Zhang W, Xu J, Liu D, Liu H, Lu X, Yu W. 2018. Characterization of an *α*-agarase from *Thalassomonas* sp. LD5 and its hydrolysate. Appl Microbiol Biotechnol: 102:2203–2212. https://doi.org/10.1007/s00253-018-8762-6.

27. Flament D, Barbeyron T, Jam M, Potin P, Czjzek M, Kloareg B, Michel G. 2007. Alpha-agarases define a new family of glycoside hydrolases, distinct from beta-agarase families. Appl Environ Microbiol 73:4691–4694. https://doi.org/10.1128/AEM.00496-07.

28. Christiansen L, Bech PK, Schultz-Johansen M, Martens HJ, Stougaard P. 2018. *Colwellia echini* sp. nov., an agar- and carrageenan-solubilizing bacterium isolated from sea urchin. Int J Syst Evol Microbiol 68:687–691. https://doi.org/10.1099/ijsem.0.002568.

29. Christiansen L, Pathiraja D, Bech PK, Schultz-Johansen M, Hennessy R, Teze D, Choi I-G, Stougaard P. 2020. A multifunctional polysaccharide utilization gene cluster in *Colwellia echini* encodes enzymes for the complete degradation of κ-carrageenan, ι-carrageenan, and hybrid *β*/κ-carrageenan. mSphere 5:e00792-19. https://doi.org/10.1128/mSphere.00792-19.

30. Henshaw J, Horne-Bitschy A, van Bueren AL, Money VA, Bolam DN, Czjzek M, Ekborg NA, Weiner RM, Hutcheson SW, Davies GJ, Boraston AB, Gilbert HJ. 2006. Family 6 carbohydrate binding modules in *β*-agarases display exquisite selectivity for the non-reducing termini of agarose chains. J Biol Chem 281:17099–17107. https://doi.org/10.1074/jbc.M600702200.

31. Kvansakul M, Adams JC, Hohenester E. 2004. Structure of a thrombospondin C-terminal fragment reveals a novel calcium core in the type 3 repeats. EMBO J 23:1223–1233. https://doi.org/10.1038/sj.emboj.7600166.

32. Hudson J, Kumar V, Egan S. 2019. Comparative genome analysis provides novel insight into the interaction of *Aquimarina* sp. AD1, BL5 and AD10 with their macroalgal host. Mar Genomics 46:8–15. https://doi.org/10.1016/j.margen.2019.02.005.

33. Ko HJ, Park E, Song J, Yang TH, Lee HJ, Kim KH, Choi IG. 2012. Functional cell surface display and controlled secretion of diverse agarolytic enzymes by *Escherichia coli* with a novel ligation-independent cloning vector based on the autotransporter YfaL. Appl Environ Microbiol 78:3051–3058. https://doi.org/10.1128/AEM.07004-11.

34. Osamu A, Inoue T, Kubo H, Minami K, Nakamura M, Iwai M, Moriyama H, Yanagisawa M, Nakasaki K. 2012. Cloning of agarase gene from non-marine agarolytic bacterium *Cellvibrio* sp. J Microbiol Biotechnol 22:1237–1244. https://doi.org/10.4014/jmb.1202.02020.

35. Ohta Y, Hatada Y, Nogi Y, Li Z, Ito S, Horikoshi K. 2004. Cloning, expression, and characterization of a glycoside hydrolase family 86 *β*-agarase from a deep-sea Microbulbifer-like isolate. Appl Microbiol Biotechnol 66:266–275. https://doi.org/10.1007/s00253-004-1757-5.

36. Frey PA. 1996. The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose. FASEB J 10:461–470. https://doi.org/10.1096/fasebj.10.4.8647345.

37. Lee SB, Cho SJ, Kim JA, Lee SY, Kim SM, Lim HS. 2014. Metabolic pathway of 3,6-anhydro-L-galactose in agar-degrading microorganisms. Biotechnol Bioproc E 19:866–878. https://doi.org/10.1007/s12257-014-0622-3.

38. Yun EJ, Lee S, Kim HT, Pelton JG, Kim S, Ko HJ, Choi IG, Kim KH. 2015. The novel catabolic pathway of 3,6-anhydro-L-galactose, the main component of red macroalgae, in a marine bacterium. Environ Microbiol 17:1677–1688. https://doi.org/10.1111/1462-2920.12607.

39. Gao C, Jin M, Yi Z, Zeng R. 2015. Characterization of a recombinant thermostable arylsulfatase from deep-sea bacterium *Flammeovirga pacifica*. J Microbiol Biotechnol 25:1894–1901. https://doi.org/10.4014/jmb.1504.04028.

40. Wang X, Duan D, Xu J, Gao X, Fu X. 2015. Characterization of a novel alkaline arylsulfatase from *Marinomonas* sp. FW-1 and its application in the desulfation of red seaweed agar. J Ind Microbiol Biotechnol 42:1353–1362. https://doi.org/10.1007/s10295-015-1625-6.

41. Reisky L, Büchsenschütz HC, Engel J, Song T, Schweder T, Hehemann J-H, Bornscheuer UT. 2018. Oxidative demethylation of algal carbohydrates by cytochrome P450 monooxygenases. Nat Chem Biol 14:342–344. https://doi.org/10.1038/s41589-018-0005-8.

42. Blanvillain S, Meyer D, Boulanger A, Lautier M, Guynet C, Denancé N, Vasse J, Lauber E, Arlat M. 2007. Plant carbohydrate scavenging through TonB-dependent receptors: a feature shared by phytopathogenic and aquatic bacteria. PLoS One 2:e224. https://doi.org/10.1371/journal.pone.0000224.

43. Thomas F, Barbeyron T, Michel G. 2011. Evaluation of reference genes for real-time quantitative PCR in the marine flavobacterium *Zobellia galactanivorans*. J Microbiol Methods 84:61–66. https://doi.org/10.1016/j.mimet.2010.10.016.

44. Neumann AM, Balmonte JP, Berger M, Giebel HA, Arnosti C, Voget S, Simon M, Brinkhoff T, Wietz M. 2015. Different utilization of alginate and other algal polysaccharides by marine *Alteromonas macleodii* ecotypes. Environ Microbiol 17:3857–3868. https://doi.org/10.1111/1462-2920.12862.

45. Yan N. 2015. Structural biology of the major facilitator superfamily transporters. Annu Rev Biophys 44:257–283. https://doi.org/10.1146/annurev-biophys-060414-033901.

46. Webb AJ, Homer KA, Hosie AH. 2008. Two closely related ABC transporters in *Streptococcus* mutans are involved in disaccharide and/or oligosaccharide uptake. J Bacteriol 190:168–178. https://doi.org/10.1128/JB.01509-07.

47. Wilson T, Ding PZ. 2001. Sodium-substrate cotransport in bacteria. Biochim Biophys Acta Bioenerg 1505:121–130. https://doi.org/10.1016/S0005-2728(00)00282-6.

48. Yu S, Yun EJ, Kim DH, Park SY, Kim KH. 2020. Dual agarolytic pathways in a marine bacterium, *Vibrio* sp. strain EJY3: molecular and enzymatic verification. Appl Environ Microbiol 86:e02724-19. https://doi.org/10.1128/AEM.02724-19.

49. Terrapon N, Lombard V, Drula E, Lapébie P, Al-Masaudi S, Gilbert HJ, Henrissat B. 2018. PULDB: the expanded database of Polysaccharide Utilization Loci. Nucleic Acids Res 46:D677–D683. https://doi.org/10.1093/nar/gkx1022.

50. Reher M, Fuhrer T, Bott M, Schönheit P. 2010. The nonphosphorylative Entner-Doudoroff pathway in the thermoacidophilic euryarchaeon *Picrophilus torridus* involves a novel 2-keto-3-deoxygluconate-specific aldolase. J Bacteriol 192:964–974. https://doi.org/10.1128/JB.01281-09.

51. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 46:W95–W101. https://doi.org/10.1093/nar/gky418.

52. Techtmann SM, Fitzgerald KS, Stelling SC, Joyner DC, Uttukar SM, Harris AP, Alshibli NK, Brown SD, Hazen TC. 2016. *Colwellia psychrerythraea* strains from distant deep-sea basins show adaptation to local conditions. Front Environ Sci 4:33. https://doi.org/10.3389/fenvs.2016.00033.

53. Gomare S, Kim HA, Ha JH, Lee MW, Park JM. 2011. Isolation of the polysaccharidase-producing bacteria from the gut of sea snail, *Batillus cornutus*. Korean J Chem Eng 28:1252–1259. https://doi.org/10.1007/s11814-010-0506-y.

54. Sawabe T, Oda Y, Shiomi Y, Ezura Y. 1995. Alginate degradation by bacteria isolated from the gut of sea urchins and abalones. Microb Ecol 30:193–202. https://doi.org/10.1007/BF00172574.

55. Pereira FC, Berry D. 2017. Microbial nutrient niches in the gut. Environ Microbiol 19:1366–1378. https://doi.org/10.1111/1462-2920.13659.

56. Fabich AJ, Jones SA, Chowdhury FZ, Cernosek A, Anderson A, Smalley D, McHargue JW, Hightower GA, Smith JT, Autieri SM, Leatham MP, Lins JJ, Allen RL, Laux DC, Cohen PS, Conway T. 2008. Comparison of carbon nutrition for pathogenic and commensal *Escherichia coli* strains in the mouse intestine. Infect Immun 76:1143–1152. https://doi.org/10.1128/IAI.01386-07.

57. Pluvinage B, Grondin JM, Amundsen C, Klassen L, Moote PE, Xiao Y, Thomas D, Pudlo NA, Anele A, Martens EC, Inglis GD, Uwiera RER, Boraston AB, Abbott DW. 2018. Molecular basis of an agarose metabolic pathway acquired by a human intestinal symbiont. Nat Commun 9:1043. https://doi.org/10.1038/s41467-018-03366-x.

58. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75.

59. Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 37:420–423. https://doi.org/10.1038/s41587-019-0036-z.

60. Juncker AS, Willenbrock H, Von Heijne G, Brunak S, Nielsen H, Krogh A. 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Sci 12:1652–1662. https://doi.org/10.1110/ps.0303703.

61. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FSL. 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26:1608–1615. https://doi.org/10.1093/bioinformatics/btq249.

62. Yu CS, Chen YC, Lu CH, Hwang JK. 2006. Prediction of protein subcellular localization. Proteins 64:643–651. https://doi.org/10.1002/prot.21018.

63. Felsenstein J. 1993. PHYLIP (phylogeny inference package), version 3.5 c. https://evolution.genetics.washington.edu/phylip.html.

64. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/molbev/msy096.

65. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.

66. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

67. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, Gottardo R, Hahne F, Hansen KD, Irizarry RA, Lawrence M, Love MI, MacDonald J, Obenchain V, Oleś AK, Pagès H, Reyes A, Shannon P, Smyth GK, Tenenbaum D, Waldron L, Morgan M. 2015. Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods 12:115–121. https://doi.org/10.1038/nmeth.3252.

68. McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res 40:4288–4297. https://doi.org/10.1093/nar/gks042.

69. Pathiraja D, Lee S, Choi I-G. 2018. Model-based complete enzymatic production of 3, 6-anhydro-L-galactose from red algal biomass. J Agric Food Chem 66:6814–6821. https://doi.org/10.1021/acs.jafc.8b01792.

70. Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 20:14. https://doi.org/10.1186/s13059-019-1832-y.

71. Kolde R. 2015. Pheatmap: Pretty Heatmaps. R package version 1.0.8. https://CRAN.R-project.org/package=pheatmap.